UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/826,162	04/15/2004	Benyu Zhang	MSI-1891US	8612

22801   7590   10/12/2006

LEE & HAYES PLLC
421 W RIVERSIDE AVENUE SUITE 500
SPOKANE, WA 99201

EXAMINER
FLEURANTIN, JEAN B

ART UNIT	PAPER NUMBER
2162	

DATE MAILED: 10/12/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

	Application No.	Applicant(s)
Office Action Summary	10/826,162	ZHANG ET AL.
	Examiner	Art Unit	
	JEAN B. FLEURANTIN	2162	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) ☒ Responsive to communication(s) filed on <u>01 April 1504</u>.
2a) ☐ This action is **FINAL**.   2b) ☒ This action is non-final.
3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) ☒ Claim(s) *1-45* is/are pending in the application.
    4a) Of the above claim(s) _____ is/are withdrawn from consideration.
5) ☐ Claim(s) _____ is/are allowed.
6) ☒ Claim(s) *1-45* is/are rejected.
7) ☐ Claim(s) _____ is/are objected to.
8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) ☒ The specification is objected to by the Examiner.
10) ☒ The drawing(s) filed on <u>15 April 2004</u> is/are: a)☒ accepted or b)☐ objected to by the Examiner.
    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
    a)☐ All   b)☐ Some * c)☐ None of:
      1.☐ Certified copies of the priority documents have been received.
      2.☐ Certified copies of the priority documents have been received in Application No. _____.
      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)
1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date <u>See Continuation Sheet</u>.
4) ☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____ .

Continuation of Attachment(s) 3). Information Disclosure Statement(s) (PTO/SB/08), Paper No(s)/Mail Date :9/18/6, 6/19/6, 1/3/6, 11/2/5, 9/9/4 and 8/2/4.

DETAILED ACTION

1.   This in response to the application filed on 04/15/04.

Claims 1-45 are presented for examination.

Information Disclosure Statement

The information disclosure statement (IDS) submitted on 08/0204, 9/09/04, 11/02/05, 6/19/06 and

9/18/06. The submission is in compliance with the provisions of 37 CFR 1.97. Accordingly, the information

disclosure statement is being considered by the examiner.

Drawings

The Drawings submitted 04/15/04 are acknowledged.

Objections

Minor informalities/Specification

Page 1, paragraph [0003], line 1, of the specification is objected to because of the missing of the

related U.S. Patent Application Number, and the specification should be amended to reflect the status of

all related application (i.e., abandoned or patented with patent number and date patented). Appropriate

correction is required.

Claim 1, lines 4 and 9, is objected to because of the terms "and/or".

Claim 6, line 5, is objected to because of the terms "and/or".

Claim 10, line 3, is objected to because of the terms "and/or".

Claim 11, line 4, is objected to because of the terms "and/or".

Claim 12, line 7, is objected to because of the terms "and/or".

Claim 13, lines 5 and 10, consecutively is objected to because of the terms "and/or".

Claim 18, line 6, is objected to because of the terms "and/or".

Claim 22, line 4 is objected to because of the terms "and/or".

Claim 25, lines 7 and 12, consecutively is objected to because of the terms "and/or".

Claim 30, line 6, is objected to because of the terms "and/or".

Claim 33, line 4, is objected to because of the terms "and/or".

Claim 34, line 5, is objected to because of the terms "and/or".

Claim 35, line 8, is objected to because of the terms "and/or".

Claim 36, lines 4 and 9 consecutively, is objected to because of the terms "and/or".

Claim 40, line 5, is objected to because of the terms "and/or".

Claim 43, line 3, is objected to because of the terms "and/or".

Claim 44, line 6, is objected to because of the terms "and/or.

Claim 45, line 8, is objected to because of the terms "and/or.

The Examiner suggests the applicant to amend the claims in order to be specific.

Claim Rejections - 35 USC § 101

35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 1-20 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

As set forth in MPEP 2106:

Products may be either machines, manufactures, or compositions of matter.
A *machine* is "a concrete thing, consisting of parts or of certain devices and combinations of devices." *Burr v. Duryee*, 68 U.S. (1 Wall.) 531, 570 (1863).

As per claim 1,

Claim 1, in view of the above cited MPEP section is not statutory, because "calculating a confidence value from combined ones of multiple similarity measurements, the combined ones comprising content, expanded, and category similarity measurements, the confidence value providing an objective measure of relevance between the bid term(s) and the site contents" does not produce any useful and tangible result.

As per claim 13,

Claim 13, in view of the above cited MPEP section is not statutory, because "calculating a confidence value from combined ones of multiple similarity measurements, the combined ones comprising content, expanded, and category similarity measurements, the confidence value providing an objective measure of relevance between the bid term(s) and the site contents" does not produce any useful and tangible result.

As per claim 25,

Claim 25, in view of the above cited MPEP section is not statutory, because "calculating a confidence value from combined ones of multiple similarity measurements, the combined ones comprising content, expanded, and category similarity measurements, the confidence value providing an objective measure of relevance between the bid term(s) and the site contents" does not produce any useful and tangible result.

As per claim 36,

Claim 36, in view of the above cited MPEP section is not statutory, because "calculating means to generate a confidence value from combined ones of multiple similarity measurements, the combined ones comprising content, expanded, and category similarity measurements, the confidence value providing an objective measure of relevance between the bid term(s) and the site contents" does not produce any useful and tangible result.

Claim 36, is directed to software per se. The claimed subject matter does not define a physical structure of hardware and software combination as required per MPEP 2106. Therefore, the claim is non-statutory

And, all dependant claims are rejected under the analysis of claims 1, 13, 25 and 36 as indicated above.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness

rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1-45 are rejected under 35 U.S.C. 103(a) as being unpatentable over "Optimizing Search

Engines using Clickthrough Data" issued to Thorsten Joachims – pages 133-142, 2002 ("Joachims"),

submitted by the Applicant(s), in view of U.S. Patent No. 6,167,398 issued to Wyard et al., ("Wyard").

As per claim 1, Joachims discloses "a method for verifying relevance between terms and Web

site contents" (i.e., method for ranking relevant documents; see page 135, col. 2, paragraph 2, lines 6-9

and equations 3-5), the method comprising:

"retrieving site contents from a bid URL" (In light the specification at paragraph [0021], the

purposed of retrieving content from a bit URL is for identifying web page. The method for clickthrough

data for identifying related queries and URLs is disclosed by Joachims page 140, col. 2, penultimate

paragraph, lines 9-10);

"formulating <u>expanded term(s)</u> semantically and/or contextually related to bid term(s)" (i.e.,

<u>maximizing</u> relative relevance; see page 135, col. 2, penultimate paragraph and equation 6),

"generating content similarity" (i.e., similarity measure; see page 135, col. 2, paragraphs 1 & 2

and equation 2) and "expanded similarity measurements from respective combinations of the bid term(s),

the site contents, and the expanded terms" (i.e., maximizing relevant documents; see page 135, col. 2,

paragraphs 2-4 and equations 2-6), "the similarity measurements indicating relatedness between

respective ones of the bid term(s), site contents, and/or expanded terms" (i.e., similarity measurement

and distance measurement; see page 135, col. 2, from line 14 to page 136, line1);

"calculating category similarity measurements between the expanded terms and the site contents in view of a similarity classifier" (i.e., using decomposition algorithms similarity, for SVM classification; see page 136, col. 2, last two paragraphs), "the combined ones comprising content, expanded, and category similarity measurements," (i.e., similarity measurement and distance measurement; see page 135, col. 2, paragraph 2 and equations 3 and 4).

Joachims fails to explicitly disclose the similarity classifier having been trained from mined web site content associated with directory data; calculating a <u>confidence value</u> from combined ones of multiple similarity measurements, the <u>confidence value</u> providing an objective measure of relevance between the bid term(s) and the site contents. However, Wyard discloses the similarity classifier having been trained from mined web site content associated with directory data (see Wyard col. 14, lines 19-25); calculating a <u>confidence value</u> from combined ones of multiple similarity measurements, the <u>confidence value</u> providing an objective measure of relevance between the bid term(s) and the site contents (see Wyard col. 10, lines 16-33). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the method of Joachims by similarity classifier having been trained from mined web site content associated with directory data; calculating a <u>confidence value</u> from combined ones of multiple similarity measurements, the <u>confidence value</u> providing an objective measure of relevance between the bid term(s) and the site contents as disclosed by Wyard (see Wyard col. 4, lines 58-65). Such a modification would allow the method of Joachims to provide a method of testing suitable of a reference document to be at least a part of a reference corpus for use in information retrieval by comparing the obtained measure with a predetermined criterion to obtain an analysis result (see Wyard col. 5, lines 45-57), therefore, improving the accuracy of the verifying relevance between keywords and web site documents.

As per claim 2, in addition to claim 1, Joachims further discloses "support vector machine (SVM)" (i.e., SVM; see page 136, col. 1, paragraph [4.1]), "a nearest neighbor (KNN)" (i.e., distance between the closest two projections; see page 136, col. 2, before last paragraph), "a decision tree" (In light the specification at paragraph [0049], a decision tree is statistical class. The method for using measure in

statistic is disclosed by Joachims at page 135, col. 2, lines 14-15). Joachims fails to explicitly disclose the similarity classifier is based on a statistical n-gram based naive Bayesian (N-Gram), a naive Bayesian (NB). However, Wyard discloses the similarity classifier is based on a statistical n-gram based naive Bayesian (N-Gram), a naive Bayesian (NB) (see Wyard col. 3, lines 36-42 and col. 4, lines 60-65 and col. 9, lines 31-32). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the method of Joachims by the similarity classifier is based on a statistical n-gram based naive Bayesian (N-Gram) as disclosed by Wyard (see Wyard col. 2, lines 32-36). Such a modification would allow the method of Joachims to provide a method of testing suitable of a reference document to be at least a part of a reference corpus for use in information retrieval by comparing the obtained measure with a predetermined criterion to obtain an analysis result (see Wyard col. 5, lines 45-57), therefore, improving the accuracy of the verifying relevance between keywords and web site documents.

As per claim 3, Joachims discloses "formulating the expanded terms further comprises generating term clusters from term vectors based on calculated term similarity" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2), "the term vectors being generated from historical queries" (i.e., weight vector match between query and document; see page 136, col. 1, paragraph [4.1] and Fig. 2), each historical query having a high frequency of occurrence" (i.e., ranking the set of documents according to a new query; see page 137, col. 1, lines 5-12), "the term clusters comprising the expanded terms" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2).

As per claim 4, in addition to claim 1, Joachims further discloses "generating respective term vectors from the bid term(s) and the site contents" (i.e., ranking queries with same featured (term, content) vectors and URLs; see page 140, col. 2, lines 27-46, particularly lines 27-36), and "calculating similarity between the respective term vectors to determine direct similarity between the bid term(s) and

the site contents" (i.e., similarity measurement and distance measurement identifying related documents and URLs; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 5, in addition to claim 1, Joachims further discloses "generating respective term vectors from the bid term(s), the site contents, and the expanded terms" (i.e., ranking queries with same featured (term, content) vectors and URLs; see page 140, col. 2, lines 27-46, particularly lines 27-36), and "calculating similarity between the respective term vectors to determine the expanded similarity measurements between the bid term(s) and the site contents" (i.e., using decomposition algorithms similar, for SVM classification; see page 136, col. 2, last two paragraphs).

As per claim 6, in addition to claim 1, Joachims further discloses "extracting features from Web site content associated with the directory data" (i.e., retrieval returning a ranking that orders documents according to their relevance query; see page 135, col. 1, paragraph [3], lines 1-5), "the features comprising a combination of title, metadata, body, hypertext link(s), visual feature(s), and/or summarization by page layout analysis information" (i.e., features number of words that query and document share inside html tags, Title; see page 136, col. 1, last two paragraphs);

"reducing dimensionality of the features via feature selection" (i.e., minimizing the number of discordant pair; see page 136, col. 1, last paragraph [3], lines 1-3 and page 137, col. 1, paragraph [4.2], lines 18-21);

categorizing the features via a classifier model to generate the similarity classifier" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2);

"generating respective term vectors from the bid term(s), the site contents, and the expanded terms" (i.e., using decomposition algorithms similarity, for SVM classification; see page 136, col. 2, last two paragraphs); and

"calculating similarity between the respective term vectors as a function of the similarity classifier to determine the category similarity measurements" (i.e., similarity measurement and distance measurement; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 7, in addition to claim 1, Joachims further discloses "training a combined relevance classifier with data of the form <term(s), Web site content, accept/reject> in view of an accept/reject threshold" (In light the specification at paragraph [0019], the purposed of training a combined relevance classifier with data in view of an accept/reject threshold is for providing different similarity measurements. The method for comparing the ordinal correlation of two random variables, Kendall's being the most frequently used measure in statistics is disclosed by Joachims page 135, col. 2, line 4 up to equation 4 and URLs col. 2, paragraph 2); and

"generating relevance verification similarity measurement (RSVM) feature vectors from the content, expanded, and category similarity measurements" (i.e., using decomposition algorithms similar, for SVM classification; see page 136, col. 2, last two paragraphs).

Joachims fails to explicitly disclose mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier. However, Wyard discloses mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier (see Wyard col. 14, lines 19-25 col. 10, lines 16-33). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the method of Joachims by mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier as disclosed by Wyard (see Wyard col. 4, lines 58-65). Such a modification would allow the method of Joachims to provide a method of testing suitable of a reference document to be at least a part of a reference corpus for use in information retrieval by comparing the obtained measure with a predetermined criterion to obtain an analysis result (see Wyard col. 5, lines 45-57), therefore, improving the accuracy of the verifying relevance between keywords and web site documents.·

As per claim 8, Joachims further discloses "caching the bid term(s) and bid URL into a bidding database" (i.e., query and the returned ranking can be easily recorded (stored); see page 134, col. 1, lines 2-4 and [caching means store/ing, according to Webster's dictionary]);

"responsive to receipt of an search query" (i.e., query result; see page 133, col. 2, paragraph [2], lines 1-3 and Fig. 1), "determining if terms of the search query are relevant to the bid term(s) in view of a possibility that the terms of the search query may not exactly match the bid term(s)" (i.e., identifying related queries and URLs mapped into features describing query and document; see page 140, col. 2, before last paragraph); and

"if the term(s) of search query are determined to be relevant to the bid term(s), communicating. the bid URL to the end-user" (i.e., ranking presenting to the user; see page 133, col. 2, paragraph [2], lines 1-6 and Fig. 1).

As per claim 9, Joachims further discloses "determining proper name similarity measurements from the bid term(s) and site contents" (i.e., vectors describing match between query and document; see page 136, col. 1, before last paragraph [4.1]),

"the proper name similarity measurements indicating relatedness between any proper name(s) detected in the bid term(s) and the site contents in view a set of proper names" (i.e., similarity measurement and distance measurement; see page 135, col. 2, paragraph 2 and equations 3 and 4); and

"wherein the combined ones of multiple similarity measurements comprise the proper name similarity measurements" (i.e., similarity measurement, distance measurement and vectors describing match between query and documents; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 10, Joachims further discloses "determining the proper name similarity measurements further comprises responsive to detecting a proper name in the bid term(s) and/or the site contents, calculating a proper name similarity score as: Prop_Sim(term, site contents), whrerein Prop_Sim(term, site contents) equals: one (1) when a term contains a proper name P, and site contents contains a conformable proper name Q; zero (0) when a term contains a proper name P, and site contents contains only unconformable proper name(s); or, zero-point-five (0.5)" (i.e., identifying related queries and URL and mapped into features describing query and document; page 140, col. 2, before last paragraph, lines 9-10).

As per claim 11, in addition to claim 1, Joachims further discloses "responsive to the determining, identifying one or more other terms that are semantically and/or contextually related to the bid URL" (i.e., identifying related queries (terms semantic) and URLs; see page 140, col. 2, before last paragraph).

As per claim 12, Joachims further discloses "generating a set of term clusters from term vectors based on calculated term similarity" (In light the specification at paragraph [0029], the purposed of generating clusters is for generating weights to group similar terms. The method for analyzing the function by studying the learned weights is disclosed by Joachims page 140, col. 1, paragraph [5.4] and Table 3), "the term vectors being generated from search engine results of submitted historical queries" (i.e., cosine matching between query and words; see page 140, col. 1, paragraph [5.4] and page 134, col. 1, lines 2-4), "each historical query having a relatively low frequency of occurrence as compared to other query terms in a query log" (i.e., largest and smallest weights; see Table 3 and page 140, col. 1, paragraph [5.4]); and "evaluating the site contents in view of term(s) specified by the term clusters to identify one or more semantically and/or contextually related terms, the terms being the one or more other terms" (In light the specification at paragraph [0029], the purposed of evaluating clusters is for generating weights to group similar terms. The method for analyzing the function by studying the learned weights is disclosed by Joachims page 140, col. 1, paragraph [5.4] and Table 3).

As per claim 13, Joachims discloses "a computer-readable medium comprising computer-executable instructions" (i.e., instructions have been executed by a computer, therefore, Joachims discloses the limitations page 140, col. 2, last paragraph, lines 10-15) "for verifying relevance between terms and Web site contents" (i.e., ranking relevant documents; see page 135, col. 2, paragraph 2, lines 6-9 and equations 3-5), the method comprising:

"retrieving site contents from a bid URL" (In light the specification at paragraph [0021], the purposed of retrieving content from a bit URL is for identifying web page. The method for clickthrough

data for identifying related queries and URL is disclosed by Joachims page 140, col. 2, penultimate paragraph, lines 9-10);

"formulating expanded term(s) semantically and/or contextually related to bid term(s)" (i.e., maximizing relative relevance see page 135, col. 2, penultimate paragraph and equation 6),

"generating content similarity" (i.e., similarity measure; see page 135, col. 2, paragraphs 1 & 2 and equation 2) and "expanded similarity measurements from respective combinations of the bid term(s), the site contents, and the expanded terms" (i.e., maximizing relevant documents; see page 135, col. 2, paragraphs 2-4 and equations 2-6), "the similarity measurements indicating relatedness between respective ones of the bid term(s), site contents, and/or expanded terms" (i.e., similarity measurement and distance measurement; see page 135, col. 2, from line 14 to page 136, line1);

"calculating category similarity measurements between the expanded terms and the site contents in view of a similarity classifier" (i.e., using decomposition algorithms similarity, for SVM classification; see page 136, col. 2, last two paragraphs), "the combined ones comprising content, expanded, and category similarity measurements," (i.e., similarity measurement and distance measurement; see page 135, col. 2, paragraph 2 and equations 3 and 4).

Joachims fails to explicitly disclose the similarity classifier having been trained from mined web site content associated with directory data; calculating a confidence value from combined ones of multiple similarity measurements, the confidence value providing an objective measure of relevance between the bid term(s) and the site contents. However, Wyard discloses the similarity classifier having been trained from mined web site content associated with directory data (see Wyard col. 14, lines 19-25); calculating a confidence value from combined ones of multiple similarity measurements, the confidence value providing an objective measure of relevance between the bid term(s) and the site contents (see Wyard col. 10, lines 16-33). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the method of Joachims by similarity classifier having been trained from mined web site content associated with directory data; calculating a confidence value from combined ones of multiple similarity measurements, the confidence value providing an objective measure of relevance between the bid term(s) and the site contents as disclosed by Wyard (see Wyard col. 4, lines

58-65): Such a modification would allow the method of Joachims to provide a method of testing suitable of a reference document to be at least a part of a reference corpus for use in information retrieval by comparing the obtained measure with a predetermined criterion to obtain an analysis result (see Wyard col. 5, lines 45-57), therefore, improving the accuracy of the verifying relevance between keywords and web site documents.

As per claim 14, in addition to claim 13, Joachims further discloses "support vector machine (SVM)" (i.e., SVM; see page 136, col. 1, paragraph [4.1]), "a nearest neighbor (KNN)" (i.e., distance between the closest two projections; see page 136, col. 2, before last paragraph), "a decision tree" (In light the specification at paragraph [0049], a decision tree is statistical class. The method for using measure in statistic is disclosed by Joachims at page 135, col. 2, lines 14-15). Joachims fails to explicitly disclose the similarity classifier is based on a statistical n-gram based naive Bayesian (N-Gram), a naive Bayesian (NB). However, Wyard discloses the similarity classifier is based on a statistical n-gram based naive Bayesian (N-Gram), a naive Bayesian (NB) (see Wyard col. 3, lines 36-42). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the method of Joachims by the similarity classifier is based on a statistical n-gram based naive Bayesian (N-Gram) as disclosed by Wyard (see Wyard col. 2, lines 32-36). Such a modification would allow the method of Joachims to provide a method of testing suitable of a reference document to be at least a part of a reference corpus for use in information retrieval by comparing the obtained measure with a predetermined criterion to obtain an analysis result (see Wyard col. 5, lines 45-57), therefore, improving the accuracy of the verifying relevance between keywords and web site documents.

As per claim 15, Joachims discloses "formulating the expanded terms further comprises generating term clusters from term vectors based on calculated term similarity" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2), "the term vectors being generated from historical queries" (i.e., weight vector match between query and document; see page 136, col. 1, paragraph [4.1] and Fig. 2), each historical query having a high frequency of

occurrence" (i.e., ranking the set of documents according to a new query; see page 137, col. 1, lines 5-12), "the term clusters comprising the expanded terms" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2).

As per claim 16, in addition to claim 13, Joachims further discloses Joachims further discloses "generating respective term vectors from the bid term(s) and the site contents" (i.e., ranking queries with same featured (term, content) vectors and URLs; see page 140, col. 2, lines 27-46, particularly lines 27-36), and "calculating similarity between the respective term vectors to determine direct similarity between the bid term(s) and the site contents" (i.e., similarity measurement and distance measurement identifying related documents and URLs; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 17, in addition to claim 13, Joachims further discloses "generating respective term vectors from the bid term(s), the site contents, and the expanded terms" (i.e., ranking queries with same featured (term, content) vectors and URLs; see page 140, col. 2, lines 27-46, particularly lines 27-36), and "calculating similarity between the respective term vectors to determine the expanded similarity measurements between the bid term(s) and the site contents" (i.e., using decomposition algorithms similar, for SVM classification; see page 136, col. 2, last two paragraphs).

As per claim 18, in addition to claim 13, Joachims further discloses "extracting features from Web site content associated with the directory data" (i.e., retrieval returning a ranking that orders documents according to their relevance query; see page 135, col. 1, paragraph [3], lines 1-5), "the features comprising a combination of title, metadata, body, hypertext link(s), visual feature(s), and/or summarization by page layout analysis information" (i.e., features number of words that query and document share inside html tags, Title; see page 136, col. 1, last two paragraphs);

"reducing dimensionality of the features via feature selection" (i.e., minimizing the number of discordant pair; see page 136, col. 1, last paragraph [3], lines 1-3 and page 137, col. 1, paragraph [4.2], lines 18-21);

categorizing the features via a classifier model to generate the similarity classifier" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2);

"generating respective term vectors from the bid term(s), the site contents, and the expanded terms" (i.e., using decomposition algorithms similarity, for SVM classification; see page 136, col. 2, last two paragraphs); and

"calculating similarity between the respective term vectors as a function of the similarity classifier to determine the category similarity measurements" (i.e., similarity measurement and distance measurement; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 19, in addition to claim 13, Joachims further discloses "training a combined relevance classifier with data of the form <term(s), Web site content, accept/reject> in view of an accept/reject threshold" (In light the specification at paragraph [0019], the purposed of training a combined relevance classifier with data in view of an accept/reject threshold is for providing different similarity measurements. The method for comparing the ordinal correlation of two random variables, Kendall's being the most frequently used measure in statistics is disclosed by Joachims page 135, col. 2, line 4 up to equation 4 and URLs col. 2, paragraph 2); and

"generating relevance verification similarity measurement (RSVM) feature vectors from the content, expanded, and category similarity measurements" (i.e., using decomposition algorithms similar, for SVM classification; see page 136, col. 2, last two paragraphs).

Joachims fails to explicitly disclose mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier. However, Wyard discloses mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier (see Wyard col. 14, lines 19-25 col. 10, lines 16-33). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the method of Joachims by mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier as disclosed by Wyard (see Wyard col. 4, lines 58-65). Such a modification would allow the method of Joachims to provide a method of testing suitable of a reference document to be at least a part of a reference corpus

for use in information retrieval by comparing the obtained measure with a predetermined criterion to obtain an analysis result (see Wyard col. 5, lines 45-57), therefore, improving the accuracy of the verifying relevance between keywords and web site documents.

As per claim 20, Joachims further discloses "caching the bid term(s) and bid URL into a bidding database" (i.e., query and the returned ranking can be easily recorded (stored); see page 134, col. 1, lines 2-4 and [caching means store/ing, according to Webster's dictionary]);

"responsive to receipt of an search query" (i.e., query result; see page 133, col. 2, paragraph [2], lines 1-3 and Fig. 1), "determining if terms of the search query are relevant to the bid term(s) in view of a possibility that the terms of the search query may not exactly match the bid term(s)" (i.e., identifying related queries and URLs mapped into features describing query and document; see page 140, col. 2, before last paragraph); and

"if the term(s) of search query are determined to be relevant to the bid term(s), communicating the bid URL to the end-user" (i.e., ranking presenting to the user; see page 133, col. 2, paragraph [2], lines 1-6 and Fig. 1).

As per claim 21, Joachims further "determining proper name similarity measurements from the bid term(s) and site contents" (i.e., vectors describing match between query and document; see page 136, col. 1, before last paragraph [4.1]),

"the proper name similarity measurements indicating relatedness between any proper name(s) detected in the bid term(s) and the site contents in view a set of proper names" (i.e., similarity measurement and distance measurement; see page 135, col. 2, paragraph 2 and equations 3 and 4); and

"wherein the combined ones of multiple similarity measurements comprise the proper name similarity measurements" (i.e., similarity measurement, distance measurement and vectors describing match between query and documents; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 22, Joachims further discloses "determining the proper name similarity measurements further comprises responsive to detecting a proper name in the bid term(s) and/or the site contents, calculating a proper name similarity score as: Prop_Sim(term, site contents), whrerein Prop_Sim(term, site contents) equals: one (1) when a term contains a proper name P, and site contents contains a conformable proper name Q; zero (0) when a term contains a proper name P, and site contents contains only unconformable proper name(s); or, zero-point-five (0.5)" (i.e., identifying related queries and URL; page 140, col. 2, before last paragraph, lines 9-10).

As per claim 23, in addition to claim 13, Joachims further discloses "responsive to the determining, identifying one or more other terms that are semantically and/or contextually related to the bid URL" (i.e., identifying related queries (terms semantic) and URLs; see page 140, col. 2, before last paragraph).

As per claim 24, Joachims further discloses "generating a set of term clusters from term vectors based on calculated term similarity" (In light the specification at paragraph [0029], the purposed of generating clusters is for generating weights to group similar terms. The method for analyzing the function by studying the learned weights is disclosed by Joachims page 140, col. 1, paragraph [5.4] and Table 3), "the term vectors being generated from search engine results of submitted historical queries" (i.e., cosine matching between query and words; see page 140, col. 1, paragraph [5.4] and page 134, col. 1, lines 2-4), "each historical query having a relatively low frequency of occurrence as compared to other query terms in a query log" (i.e., largest and smallest weights; see Table 3 and page 140, col. 1, paragraph [5.4]); and "evaluating the site contents in view of term(s) specified by the term clusters to identify one or more semantically and/or contextually related terms, the terms being the one or more other terms" (In light the specification at paragraph [0029], the purposed of evaluating clusters is for generating weights to group similar terms. The method for analyzing the function by studying the learned weights is disclosed by Joachims page 140, col. 1, paragraph [5.4] and Table 3).

As per claim 25, Joachims discloses "a computing device for verifying relevance between terms and Web site contents" (i.e., ranking relevant documents; see page 135, col. 2, paragraph 2, lines 6-9 and equations 3-5), the computing comprising:

"a processor; and a memory coupled to the processor, the memory comprising computer-program instructions executable by the processor" (i.e., instructions have been executed by a computer, therefore, Joachims discloses the limitations page 140, col. 2, last paragraph, lines 10-15) for:

"retrieving site contents from a bid URL" (In light the specification at paragraph [0021], the purposed of retrieving content from a bit URL is for identifying web page. The method for clickthrough data for identifying related queries and URL is disclosed by Joachims page 140, col. 2, penultimate paragraph, lines 9-10);

"formulating expanded term(s) semantically and/or contextually related to bid term(s)" (i.e., maximizing relative relevance; see page 135, col. 2, penultimate paragraph and equation 6),

"generating content similarity" (i.e., similarity measure; see page 135, col. 2, paragraphs 1 & 2 and equation 2) and "expanded similarity measurements from respective combinations of the bid term(s), the site contents, and the expanded terms" (i.e., maximizing relevant documents; see page 135, col. 2, paragraphs 2-4 and equations 2-6), "the similarity measurements indicating relatedness between respective ones of the bid term(s), site contents, and/or expanded terms" (i.e., similarity measurement and distance measurement; see page 135, col. 2, from line 14 to page 136, line1);

"calculating category similarity measurements between the expanded terms and the site contents in view of a similarity classifier" (i.e., using decomposition algorithms similarity, for SVM classification; see page 136, col. 2, last two paragraphs), "the combined ones comprising content, expanded, and category similarity measurements," (i.e., similarity measurement and distance measurement; see page 135, col. 2, paragraph 2 and equations 3 and 4).

Joachims fails to explicitly disclose the similarity classifier having been trained from mined web site content associated with directory data; calculating a confidence value from combined ones of multiple

similarity measurements, the confidence value providing an objective measure of relevance between the bid term(s) and the site contents. However, Wyard discloses the similarity classifier having been trained from mined web site content associated with directory data (see Wyard col. 14, lines 19-25); calculating a confidence value from combined ones of multiple similarity measurements, the confidence value providing an objective measure of relevance between the bid term(s) and the site contents (see Wyard col. 10, lines 16-33). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the method of Joachims by similarity classifier having been trained from mined web site content associated with directory data; calculating a confidence value from combined ones of multiple similarity measurements, the confidence value providing an objective measure of relevance between the bid term(s) and the site contents as disclosed by Wyard (see Wyard col. 4, lines 58-65). Such a modification would allow the method of Joachims to provide a method of testing suitable of a reference document to be at least a part of a reference corpus for use in information retrieval by comparing the obtained measure with a predetermined criterion to obtain an analysis result (see Wyard col. 5, lines 45-57), therefore, improving the accuracy of the verifying relevance between keywords and web site documents.

As per claim 26, in addition to claim 25, Joachims further discloses "support vector machine (SVM)" (i.e., SVM; see page 136, col. 1, paragraph [4.1]), "a nearest neighbor (KNN)" (i.e., distance between the closest two projections; see page 136, col. 2, before last paragraph), "a decision tree" (In light the specification at paragraph [0049], a decision tree is statistical class. The method for using measure in statistic is disclosed by Joachims at page 135, col. 2, lines 14-15). Joachims fails to explicitly disclose the similarity classifier is based on a statistical n-gram based naive Bayesian (N-Gram), a naive Bayesian (NB). However, Wyard discloses the similarity classifier is based on a statistical n-gram based naive Bayesian (N-Gram), a naive Bayesian (NB) (see Wyard col. 3, lines 36-42). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the method of Joachims by the similarity classifier is based on a statistical n-gram based naive Bayesian (N-Gram) as disclosed by Wyard (see Wyard col. 2, lines 32-36). Such a modification would allow the method of

Joachims to provide a method of testing suitable of a reference document to be at least a part of a reference corpus for use in information retrieval by comparing the obtained measure with a predetermined criterion to obtain an analysis result (see Wyard col. 5, lines 45-57), therefore, improving the accuracy of the verifying relevance between keywords and web site documents.

As per claim 27, Joachims discloses "formulating the expanded terms further comprises generating term clusters from term vectors based on calculated term similarity" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2), "the term vectors being generated from historical queries" (i.e., weight vector match between query and document; see page 136, col. 1, paragraph [4.1] and Fig. 2), each historical query having a high frequency of occurrence" (i.e., ranking the set of documents according to a new query; see page 137, col. 1, lines 5-12), "the term clusters comprising the expanded terms" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2).

As per claim 28, in addition to claim 25, Joachims further discloses Joachims further discloses "generating respective term vectors from the bid term(s) and the site contents" (i.e., ranking queries with same featured (term, content) vectors and URLs; see page 140, col. 2, lines 27-46, particularly lines 27-36), and "calculating similarity between the respective term vectors to determine direct similarity between the bid term(s) and the site contents" (i.e., similarity measurement and distance measurement identifying related documents and URLs; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 29, in addition to claim 25, Joachims further discloses "generating respective term vectors from the bid term(s), the site contents, and the expanded terms" (i.e., ranking queries with same featured (term, content) vectors and URLs; see page 140, col. 2, lines 27-46, particularly lines 27-36), and "calculating similarity between the respective term vectors to determine the expanded similarity measurements between the bid term(s) and the site contents" (i.e., using decomposition algorithms similar, for SVM classification; see page 136, col. 2, last two paragraphs).

As per claim 30, in addition to claim 25, Joachims further discloses "extracting features from Web site content associated with the directory data" (i.e., retrieval returning a ranking that orders documents according to their relevance query; see page 135, col. 1, paragraph [3], lines 1-5), "the features comprising a combination of title, metadata, body, hypertext link(s), visual feature(s), and/or summarization by page layout analysis information" (i.e., features number of words that query and document share inside html tags, Title; see page 136, col. 1, last two paragraphs);

"reducing dimensionality of the features via feature selection" (i.e., minimizing the number of discordant pair; see page 136, col. 1, last paragraph [3], lines 1-3 and page 137, col. 1, paragraph [4.2], lines 18-21);

categorizing the features via a classifier model to generate the similarity classifier" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2);

"generating respective term vectors from the bid term(s), the site contents, and the expanded terms" (i.e., using decomposition algorithms similarity, for SVM classification; see page 136, col. 2, last two paragraphs); and

"calculating similarity between the respective term vectors as a function of the similarity classifier to determine the category similarity measurements" (i.e., similarity measurement and distance measurement; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 31, in addition to claim 25, Joachims further discloses "training a combined relevance classifier with data of the form <term(s), Web site content, accept/reject> in view of an accept/reject threshold" (In light the specification at paragraph [0019], the purposed of training a combined relevance classifier with data in view of an accept/reject threshold is for providing different similarity measurements. The method for comparing the ordinal correlation of two random variables, Kendall's being the most frequently used measure in statistics is disclosed by Joachims page 135, col. 2, line 4 up to equation 4 and URLs col. 2, paragraph 2); and

"generating relevance verification similarity measurement (RSVM) feature vectors from the content, expanded, and category similarity measurements" (i.e., using decomposition algorithms similar, for SVM classification; see page 136, col. 2, last two paragraphs).

Joachims fails to explicitly disclose mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier. However, Wyard discloses mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier (see Wyard col. 14, lines 19-25 col. 10, lines 16-33). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the method of Joachims by mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier as disclosed by Wyard (see Wyard col. 4, lines 58-65). Such a modification would allow the method of Joachims to provide a method of testing suitable of a reference document to be at least a part of a reference corpus for use in information retrieval by comparing the obtained measure with a predetermined criterion to obtain an analysis result (see Wyard col. 5, lines 45-57), therefore, improving the accuracy of the verifying relevance between keywords and web site documents.

As per claim 32, Joachims further discloses "determining proper name similarity measurements from the bid term(s) and site contents" (i.e., vectors describing match between query and document; see page 136, col. 1, before last paragraph [4.1]),

"the proper name similarity measurements indicating relatedness between any proper name(s) detected in the bid term(s) and the site contents in view a set of proper names" (i.e., similarity measurement and distance measurement; see page 135, col. 2, paragraph 2 and equations 3 and 4); and

"wherein the combined ones of multiple similarity measurements comprise the proper name similarity measurements" (i.e., similarity measurement, distance measurement and vectors describing match between query and documents; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 33, Joachims further discloses "determining the proper name similarity measurements further comprises responsive to detecting a proper name in the bid term(s) and/or the site contents, calculating a proper name similarity score as: Prop_Sim(term, site contents), whrerein Prop_Sim(term, site contents) equals: one (1) when a term contains a proper name P, and site contents contains a conformable proper name Q; zero (0) when a term contains a proper name P, and site contents contains only unconformable proper name(s); or, zero-point-five (0.5)" (i.e., identifying related queries and URL; page 140, col. 2, before last paragraph, lines 9-10).

As per claim 34, in addition to claim 25, Joachims further discloses "responsive to the determining, identifying one or more other terms that are semantically and/or contextually related to the bid URL" (i.e., identifying related queries (terms semantic) and URLs; see page 140, col. 2, before last paragraph).

As per claim 35, Joachims further discloses "generating a set of term clusters from term vectors based on calculated term similarity" (In light the specification at paragraph [0029], the purposed of generating clusters is for generating weights to group similar terms. The method for analyzing the function by studying the learned weights is disclosed by Joachims page 140, col. 1, paragraph [5.4] and Table 3), "the term vectors being generated from search engine results of submitted historical queries" (i.e., cosine matching between query and words; see page 140, col. 1, paragraph [5.4] and page 134, col. 1, lines 2-4), "each historical query having a relatively low frequency of occurrence as compared to other query terms in a query log" (i.e., largest and smallest weights; see Table 3 and page 140, col. 1, paragraph [5.4]); and "evaluating the site contents in view of term(s) specified by the term clusters to identify one or more semantically and/or contextually related terms, the terms being the one or more other terms" (In light the specification at paragraph [0029], the purposed of evaluating clusters is for generating weights to group similar terms. The method for analyzing the function by studying the learned weights is disclosed by Joachims page 140, col. 1, paragraph [5.4] and Table 3).

As per claim 36, Joachims discloses "a computing device for verifying relevance between terms and Web site contents" (i.e., method for ranking relevant documents; see page 135, col. 2, paragraph 2, lines 6-9 and equations 3-5), the computing device comprising:

"retrieving means to obtain site contents from a bid URL" (In light the specification at paragraph [0021], the purposed of retrieving content from a bit URL is for identifying web page. The method for clickthrough data for identifying related queries and URL is disclosed by Joachims page 140, col. 2, penultimate paragraph, lines 9-10);

"formulating means to identify <u>expanded term(s)</u> semantically and/or contextually related to bid term(s)" (i.e., <u>maximizing</u> relative relevance; see page 135, col. 2, penultimate paragraph and equation 6),

"generating means to create content similarity" (i.e., similarity measure; see page 135, col. 2, paragraph 2 and equations 1 & 2) and "expanded similarity measurements from respective combinations of the bid term(s), the site contents, and the expanded terms" (i.e., maximizing relevant documents; see page 135, col. 2, paragraphs 2-4 and equations 2-6), "the similarity measurements indicating relatedness between respective ones of the bid term(s), site contents, and/or expanded terms" (i.e., similarity measurement and distance measurement; see page 135, col. 2, from line 14 to page 136, line1);

"calculating means to determine category similarity measurements between the expanded terms and the site contents in view of a similarity classifier" (i.e., using decomposition algorithms similarity, for SVM classification; see page 136, col. 2, last two paragraphs), "the combined ones comprising content, expanded, and category similarity measurements," (i.e., similarity measurement and distance measurement; see page 135, col. 2, paragraph 2 and equations 3 and 4).

Joachims fails to explicitly disclose the similarity classifier having been trained from mined web site content associated with directory data; calculating means to generate a <u>confidence value</u> from combined ones of multiple similarity measurements, the <u>confidence value</u> providing an objective measure of relevance between the bid term(s) and the site contents. However, Wyard discloses disclose the similarity classifier having been trained from mined web site content associated with directory data; calculating means to generate a <u>confidence value</u> from combined ones of multiple similarity

measurements, the <u>confidence value</u> providing an objective measure of relevance between the bid term(s) and the site contents (see Wyard col. 10, lines 16-33). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the method of Joachims by disclose the similarity classifier having been trained from mined web site content associated with directory data; calculating means to generate a <u>confidence value</u> from combined ones of multiple similarity measurements, the <u>confidence value</u> providing an objective measure of relevance between the bid term(s) and the site contents as disclosed by Wyard (see Wyard col. 4, lines 58-65). Such a modification would allow the method of Joachims to provide a method of testing suitable of a reference document to be at least a part of a reference corpus for use in information retrieval by comparing the obtained measure with a predetermined criterion to obtain an analysis result (see Wyard col. 5, lines 45-57), therefore, improving the accuracy of the verifying relevance between keywords and web site documents.

As per claim 37, Joachims discloses "the computer formulating means further comprise to generating to create term clusters from term vectors based on calculated term similarity" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2), "the term vectors being generated from historical queries" (i.e., weight vector match between query and document; see page 136, col. 1, paragraph [4.1] and Fig. 2), each historical query having a high frequency of occurrence" (i.e., ranking the set of documents according to a new query; see page 137, col. 1, lines 5-12), "the term clusters comprising the expanded terms" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2).

As per claim 38, in addition to claim 36, Joachims further discloses Joachims further discloses "generating respective term vectors from the bid term(s) and the site contents" (i.e., ranking queries with same featured (term, content) vectors and URLs; see page 140, col. 2, lines 27-46, particularly lines 27-36), and "calculating similarity between the respective term vectors to determine direct similarity between the bid term(s) and the site contents" (i.e., similarity measurement and distance measurement identifying related documents and URLs; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 39, in addition to claim 36, Joachims further discloses "creating means to generate respective term vectors from the bid term(s), the site contents, and the expanded terms" (i.e., ranking queries with same features (term, content) vectors and URLs; see page 140, col. 2, lines 27-46, particularly lines 27-36), and

"calculating means to determine similarity between the respective term vectors to determine the expanded similarity measurements between the bid term(s) and the site contents" (i.e., using decomposition algorithms similar, for SVM classification; see page 136, col. 2, last two paragraphs).

As per claim 40, in addition to claim 36, Joachims further discloses "extracting means to obtain features from Web site content associated with the directory data" (i.e., retrieval returning a ranking that orders documents according to their relevance query; see page 135, col. 1, paragraph [3], lines 1-5), "the features comprising a combination of title, metadata, body, hypertext link(s), visual feature(s), and/or summarization by page layout analysis information" (i.e., features number of words that query and document share inside html tags, Title; see page 136, col. 1, last two paragraphs);

"reducing means to lessen dimensionality of the features via feature selection" (i.e., minimizing the number of discordant pair; see page 136, col. 1, last paragraph [3], lines 1-3 and page 137, col. 1, paragraph [4.2], lines 18-21);

categorizing means to organize the features via a classifier model to generate the similarity classifier" (i.e., clustering using homogenous groups; see page 141, col. 1, 'section conclusion and future work' - paragraph 2);

"generating means to create respective term vectors from the bid term(s), the site contents, and the expanded terms" (i.e., using decomposition algorithms similarity, for SVM classification; see page 136, col. 2, last two paragraphs); and

"calculating means to identify similarity between the respective term vectors as a function of the similarity classifier to determine the category similarity measurements" (i.e., identifying similarity measurement and distance measurement; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 41, in addition to claim 36, Joachims further discloses "training a combined relevance classifier with data of the form <term(s), Web site content, accept/reject> in view of an accept/reject threshold" (In light the specification at paragraph [0019], the purposed of training a combined relevance classifier with data in view of an accept/reject threshold is for providing different similarity measurements. The method for comparing the ordinal correlation of two random variables, Kendall's being the most frequently used measure in statistics is disclosed by Joachims page 135, col. 2, line 4 up to equation 4 and URLs col. 2, paragraph 2); and

"generating relevance verification similarity measurement (RSVM) feature vectors from the content, expanded, and category similarity measurements" (i.e., using decomposition algorithms similar, for SVM classification; see page 136, col. 2, last two paragraphs).

Joachims fails to explicitly disclose mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier. However, Wyard discloses mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier (see Wyard col. 14, lines 19-25 col. 10, lines 16-33). It would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the method of Joachims by mapping multiple scores from the RSVM feature vectors to the confidence value via the combined relevance classifier as disclosed by Wyard (see Wyard col. 4, lines 58-65). Such a modification would allow the method of Joachims to provide a method of testing suitable of a reference document to be at least a part of a reference corpus for use in information retrieval by comparing the obtained measure with a predetermined criterion to obtain an analysis result (see Wyard col. 5, lines 45-57), therefore, improving the accuracy of the verifying relevance between keywords and web site documents.

As per claim 42, Joachims further discloses "determining means to determine proper name similarity measurements from the bid term(s) and site contents" (i.e., vectors describing match between query and document; see page 136, col. 1, before last paragraph [4.1]),

"the proper name similarity measurements indicating relatedness between any proper name(s) detected in the bid term(s) and the site contents in view a set of proper names" (i.e., similarity measurement and distance measurement; see page 135, col. 2, paragraph 2 and equations 3 and 4); and

"wherein the combined ones of multiple similarity measurements comprise the proper name similarity measurements" (i.e., similarity measurement, distance measurement and vectors describing match between query and documents; see page 135, col. 2, from line 14 to page 136, line1).

As per claim 43, Joachims further discloses "determining means to determine the proper name similarity measurements further comprises responsive to detecting a proper name in the bid term(s) and/or the site contents, calculating a proper name similarity score" (i.e., identifying related queries and URL; page 140, col. 2, before last paragraph, lines 9-10).

As per claim 44, in addition to claim 36, Joachims further discloses "responsive to the determining, identifying means to identify one or more other terms that are semantically and/or contextually related to the bid URL" (i.e., identifying related queries (terms semantic) and URLs; see page 140, col. 2, before last paragraph).

As per claim 45, Joachims further discloses "generating means to generate a set of term clusters from term vectors based on calculated term similarity" (In light the specification at paragraph [0029], the purposed of generating clusters is for generating weights to group similar terms. The method for analyzing the function by studying the learned weights is disclosed by Joachims page 140, col. 1, paragraph [5.4] and Table 3), "the term vectors being generated from search engine results of submitted historical queries" (i.e., cosine matching between query and words; see page 140, col. 1, paragraph [5.4] and page 134, col. 1, lines 2-4), "each historical query having a relatively low frequency of occurrence as compared to other query terms in a query log" (i.e., largest and smallest weights; see Table 3 and page 140, col. 1, paragraph [5.4]); and

"evaluating means to evaluate the site contents in view of term(s) specified by the term clusters to identify one or more semantically and/or contextually related terms, the terms being the one or more other terms" (In light the specification at paragraph [0029], the purposed of evaluating clusters is for generating weights to group similar terms. The method for analyzing the function by studying the learned weights is disclosed by Joachims page 140, col. 1, paragraph [5.4] and Table 3).

Prior Art

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Sharon McDonald et al., Evaluating a Content Based Image Retrieval System.

Prager, U.S. Patent No. 6,003,027 relates to determining the confidence of results from categorization systems.

CONTACT INFORMATION

2.   Any inquiry concerning this communication or earlier communications from the examiner should be directed to JEAN B. FLEURANTIN whose telephone number is 571 – 272-4035. The examiner can normally be reached on 7:05 to 4:35.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, JOHN E BREENE can be reached on 571 – 272-4107. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Jean Bolte Fleurantin

Patent Examiner

Technology Center 2100

October 04, 2006